

1. Загальна інформація	
Назва дисципліни	Big Data
Рівень вищої освіти	Перший (бакалаврський) рівень
Викладач	Миронова Наталя Олексіївна, канд. техн. наук., доцент
Контактна інформація викладача	natali.myronova@gmail.com
Час і місце проведення навчальної дисципліни	<i>Предметна аудиторія кафедри</i>
Обсяг дисципліни	Кількість кредитів – 4. Загальний обсяг – 120 год. Модулів – 2. Лекції – 30 год. (денна). Лабораторні – 14 год. (денна). Самостійна – 76 год. (денна). Вид контролю – залік. Мова викладання може бути як українська, так і англійська (за попередньою домовленістю зі студентами)
Консультації	Згідно з графіком консультацій, дистанційно
2. Пререквізити і постреквізити навчальної дисципліни	
<p>Перелік дисциплін, вивчення яких має передувати дисципліні:</p> <ul style="list-style-type: none"> – Вища математика. – Програмування. – Системи керування базами даних. – Технології програмування. <p>Перелік дисциплін, для вивчення яких є обов'язковими знання, здобуті при вивченні цієї дисципліни:</p> <ul style="list-style-type: none"> – Виконання розділу <i>дипломного проекту</i>. 	
3. Характеристика навчальної дисципліни	
<p>Дисципліна «Big Data» є вибірковою компонентою у підготовці студентів освітньо-професійної програми «Інтелектуальні мехатронні та робототехнічні системи». Мета навчальної дисципліни - розглянути сучасні методи та програмні засоби обробки структурованих та неструктурованих великих даних. Ознайомлення та отримання практичних навиків з аналізу та обробки великих даних з використанням сучасних технологій.</p> <p>Перелік компетентностей, яких набуває студент при вивченні.</p> <p>Загальні компетентності:</p> <p>K01. Здатність застосовувати знання у практичних ситуаціях. K02. Здатність спілкуватися державною мовою як усно, так і письмово. K03. Здатність спілкуватися іноземною мовою. K04. Навички використання інформаційних і комунікаційних технологій. K05. Здатність до пошуку, опрацювання та аналізу інформації з різних джерел. K08. Здатність працювати в команді.</p> <p>Фахові компетентності:</p> <p>K11. Здатність застосовувати знання математики, в обсязі, необхідному для використання математичних методів для аналізу і синтезу систем автоматизації. K16. Здатність використовувати для вирішення професійних завдань новітні технології у галузі автоматизації та комп'ютерно-інтегрованих технологій, зокрема, проектування багаторівневих систем керування, збору даних та їх архівування для формування бази даних параметрів процесу та їх візуалізації за допомогою засобів людино-машинного інтерфейсу. K19. Здатність вільно користуватись сучасними комп'ютерними та інформаційними технологіями для вирішення професійних завдань, програмувати та використовувати</p>	

прикладні та спеціалізовані комп'ютерно-інтегровані середовища для вирішення задач автоматизації.

Очікувані програмні результати навчання:

ПР03. Вміти застосовувати сучасні інформаційні технології та мати навички розробляти алгоритми та комп'ютерні програми з використанням мов високого рівня та технологій об'єктно-орієнтованого програмування, створювати бази даних та використовувати інтернет-ресурси.

ПР09. Вміти проектувати багаторівневі системи керування і збору даних для формування бази параметрів процесу та їх візуалізації за допомогою засобів людино-машинного інтерфейсу, використовуючи новітні комп'ютерно-інтегровані технології.

ПР012. Вміти використовувати різноманітне спеціалізоване програмне забезпечення для розв'язування типових інженерних задач у галузі автоматизації, зокрема, математичного моделювання, автоматизованого проектування, керування базами даних, методів комп'ютерної графіки.

4. Мета вивчення навчальної дисципліни

Грунтовне ознайомлення студентів з сучасними програмними засобами та новітніми комп'ютерно-інтегрованими технологіями для аналізу и обробки Big Data

5. Завдання вивчення дисципліни

Головним завданням вивчення курсу є навчити студентів користуватися методами обробки структурованих та неструктурованих даних, використовуючи різні сучасні програмні засоби та новітні комп'ютерно-інтегровані технології

6. Зміст навчальної дисципліни

Структура навчальної дисципліни:

1. Введення до Big Data
2. Програмне забезпечення обробки великих даних Apache Hadoop.
3. Програмне забезпечення обробки великих даних Apache Spark.
4. Вирішення прикладних практичних завдань.

7. План вивчення навчальної дисципліни

№ тижня	Назва теми	Форми організації навчання	Кількість годин
1	Введення до Big Data.	Лекції Лаб.роботи Сам. робота	2 - 2
2	Основні поняття та визначення аналізу та обробки великих даних. Лабораторна робота №1 Розгортання віртуальної машини з програмним забезпечення обробки великих даних Apache Hadoop	Лекції Лаб.роботи Сам. робота	2 2 6
3	Введення до екосистеми Apache Hadoop.	Лекції Лаб.роботи Сам. робота	2 - 6
4	Введення до Hadoop Distributed File System та MapReduce Лабораторна робота №2 Вивчення компонентів ядра Apache Hadoop: Hadoop Distributed File System та MapReduce	Лекції Лаб.роботи Сам. робота	2 2 8
5	Введення до Hadoop Streaming. Лабораторна робота №3 Вивчення принципів роботи Hadoop Streaming	Лекції Лаб.роботи Сам. робота	2 2 4
6	Введення до до екосистеми Apache Spark	Лекції Лаб.роботи Сам. робота	2 - 8
7	Поняття Resilient Distributed Dataset Лабораторна робота №4 Основи роботи з	Лекції Лаб.роботи	2 2

	Resilient Distributed Dataset	Сам. робота	8
8	Поняття DataFrame Лабораторна робота №5 Основи роботи з DataFrame	Лекції Лаб.роботи Сам. робота	2 2 8
9	Введення до Spark SQL	Лекції Лаб.роботи Сам. робота	2 - 4
10-11	Введення до Machine learning Spark library Лабораторна робота №6 Основи роботи з Machine learning Spark library	Лекції Лаб.роботи Сам. робота	4 2 8
12-15	Огляд практичних прикладних завдань Лабораторна робота №7 Вирішення прикладних завдань	Лекції Лаб.роботи Сам. робота	8 2 14

8. Самостійна робота

Назва теми		Кількість годин
1	Історія розвитку Big Data.	2
2	Історія розвитку програмного забезпечення обробки великих даних Apache Hadoop. Огляд дистрибутивів та хмарних сервісів Apache Hadoop.	6
3	Огляд компонентів Apache Hadoop: HDFS, MapReduce, YARN, HBase, Hive, Pig, Sqoop, Flume, Kafka, ZooKeeper, Spark	6
4	Парадигма програмування MapReduce.	8
5	Програмування для Hadoop Streaming з використанням мови Python	4
6	Порівняльна характеристика Apache Hadoop та Apache Spark. Компоненти Apache Spark. Функціональне програмування з використанням мови Python. Лямбда функції, загальні функціональні функції, структури даних.	8
7	Операції над Resilient Distributed Dataset.	8
8	Операції над DataFrame.	8
9	Вивчення основних функцій Spark SQL	4
10	Вивчення основних функцій Machine learning Spark library	8
11	Вирішення практичних прикладних завдань	14
Разом		76

9. Система та критерії оцінювання курсу

Успішність студентів денної форми навчання оцінюється за результатами:

- опитування з кожної теми;
- захисту звітів про виконання лабораторних робіт;
- рубіжного контролю за кожний блок змістовних модулів;
- захисту звітів про виконання індивідуальних завдань;
- рубіжних модульних контролів за кожний блок змістовних модулів.

Розподіл балів, які отримують студенти

Поточне оцінювання (аудиторна та самостійна робота)								Кількість балів (залік)	Сумарна к-ть балів
Змістовий модуль 1				Змістовий модуль 2				40	100
ЛР1	ЛР2	ЛР3	ЛР4	ЛР5	ЛР6	ЛР7	Тест		
5	10	5	5	5	10	10	10		

10. Політика курсу

Індивідуальні завдання: Студенти, що бажають заробити додаткові бали (до 20), можуть самостійно зареєструватися на курсах платформ Coursera або Udemu, попередньо узгодивши тематику обраного курсу або курсів, повинні отримати відповідний сертифікат або сертифікати і показати його (їх) викладачу. Кількість балів буде виставлена пропорційно до успіхів студента (досягнення на курсі згідно зі статистикою Coursera або Udemu, сумарна мінімальна кількість годин курсу або курсів - 30 або 1 ECTS). Також можлива участь студентів в міжнародних школах та семінарах з Big Data, студенти повинні отримати відповідний сертифікат та показати його викладачу (мінімальна кількість годин 30 або 1 ECTS).

Написання та публікація тез доповіді (одних з дисципліни) на науково-практичну конференцію викладачів, науковців, молодих учених, аспірантів та студентів «Тиждень науки» оцінюється в 10 додаткових балів. Написання та публікація тез доповіді (одних з дисципліни) на міжнародну конференцію або однієї спільної з викладачем наукової статі, що реферуються в SCOPUS, оцінюється в 20 додаткових балів.

Політика щодо крайніх термінів складання та перескладання:

Роботи, які здаються із порушенням термінів без поважних причин, оцінюються на нижчу оцінку (60% від можливої максимальної кількості балів за вид діяльності балів). Перескладання модулів відбувається за наявності поважних причин (наприклад, лікарняний).

Політика щодо академічної доброчесності: Списування під час контрольних робіт заборонені (в тому числі із використанням мобільних пристроїв). Мобільні пристрої дозволяється використовувати лише під час он-лайн тестування та підготовки лабораторних завдань під час заняття.

Політика щодо відвідування: Відвідування занять є обов'язковим компонентом оцінювання. За об'єктивних причин (наприклад, хвороба, працевлаштування, міжнародне стажування) навчання може відбуватись в онлайн формі за погодженням із керівником курсу.